DOCUMENT RESUME

ED 289 869 SP 029 786

TITLE	Teacher Assessment.
INSTITUTION	Association of Teacher Educators, Reston, Va.
PUB DATE	88
NOTE	63p.
AVAILABLE FROM	Association of Teacher Educators, 1900 Association Drive, Reston, VA 22091 ($5.50).
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Educational Research; *Evaluation Methods; Legal Problems; *Policy Formation; Political Issues; Teacher Effectiveness; *Teacher Evaluation; Testing

ABSTRACT

This monograph explores many of the issues related to teacher assessment. In Section I, the criteria for teacher assessment and evaluation are discussed. An analysis is given of the issues involved in teacher assessment in the second section. Among the issues considered are the validity of research findings on teacher effectiveness, measurement techniques, the political implications of teacher evaluation, and legal issues involved in occupational assessments. The question of using evaluation to improve teaching is also discussed. Recommendations are made on dealing with the issues involved in teacher evaluation, and 42 references are listed. The appendix contains a list of state-mandated tests for prospective teachers and their uses and purposes. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED289869

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

R. Stevenson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SP 029 786

Teacher Assessment

2

Association of Teacher Educators

Officers

President—Janet Towslee Collier, Dean, College of Education, Georgia State University, Atlanta, Georgia

President-Elect—Bill Dixon, Chair, Department of Curriculum and Instruction, College of Education, Southern Illinois University, Carbondale, Illinois

Past President—Robert F. Schuck, Dean, College of Education and Human Development, Rhode Island College, Providence, Rhode Island

Executive Board Members

Jean L. Easterly, Chair, Teacher Education, California State University, Hayward, California

B. Keith Eicher, Coordinator of Field Experiences, University of Richmond, Richmond, Virginia

Stuart Ervay, Director, Center for Educational Research and Service, Emporia State University, Emporia, Kansas

Howard D. Hill, Director of Chapter Programs, Phi Delta Kappa International, Bloomington, Indiana

D. John McIntyre, Director, Teaching Skills Laboratory, and Associate Professor, Curriculum, Instruction, and Media, Southern Illinois University, Carbondale, Illinois

Lyle Mottinger, Principal, Hosterman Junior High, New Hope, Minnesota

Dora Scott, Teacher, Sutton Elementary School, Houston, Texas

Executive Director

Robert J. Stevenson

Commission on Teacher Assessment

Theodore E. Andrews (Chair), Director of Professional Education, Superintendent of Public Instruction, Olympia, Washington

Susan Barnes, Director, Division of Teacher Education, Texas Education Agency, Austin, Texas

Donald R. Cruickshank, Professor, Educational Foundations and Research, College of Education, The Ohio State University, Columbus, Ohio

Roy A. Edelfelt, Educational Consultant, Edelfelt Johnson, Chapel Hill, North Carolina

Donald Haefele, Associate Professor, Educational Services and Research, College of Education, The Ohio State University, Columbus, Ohio

John H. Hansen, Professor, College of Education, Florida State University, Tallahassee, Florida

Ray Kurtz, Professor, Department of Curriculum and Instruction, College of Education, Kansas State University, Manhattan, Kansas

Richard Peterson, Senior Research Psychologist, Educational Testing Service, Berkeley, California

Hersh Waxman, Director, Educational Research Center, University of Houston, Houston, Texas

Teacher Assessment

Teacher Assessment

Association of Teacher Educators
Reston, Virginia

Association of Teacher Educators
1900 Association Drive
Reston, VA 22091

© 1988 by the Association of Teacher Educators
All rights reserved
Printed in the United States of America

Cover Design	Susan Hankoff-Estrella Alexandria, Virginia
Editing, Text Design, and Composition	Edelfelt Johnson Chapel Hill, North Carolina

Library of Congress Cataloging in Publication Data

Association of Teacher Educators. Commission on Teacher Assessment.
Teacher assessment.
p. cm.
Bibliography: p.
1. Teachers—United States—Rating of. 2. Prediction of teacher success. I. Title.
LB2838.A83 1988
371.1'44—dc19 87-33447
CIP

Contents

Foreword	*v*
Preface	*vii*
Introduction	*1*
The Commission's Convictions About Teacher Assessment and Evaluation	*7*
Issues in Assessment	*11*
Research Issues	*12*
Measurement Issues	*16*
Political Issues	*20*
Legal Issues	*26*
An Often-Neglected Issue: Evaluation to Improve Teaching	*32*
Recommendations	*41*
References	*45*
Appendix: State-Mandated Testing of Prospective Teachers: Test Used and Purpose	*51*

Foreword

The eagerness to ensure better education, which means ensuring quality teaching, is a goal of the 1980s with parents, citizens, government officials, and elected representatives. The desire is healthy. Attention to schools and teachers is something most educators have wanted for a long time. The action so far has included a hard look at teaching—how to define it, how to measure it, and how to nurture it. The tough part is how to assess it.

The motivation for assessment is positive. It can be productive, but only if done with care, intelligence, and fairness. Ensuring quality in education may be the first or second human priority after survival. After all, the next generation, the future of the country, is at stake.

There are always fits and starts with reform efforts. Assessing teachers is no exception. Some forward steps have been taken, and some mistakes have been made. This publication is an effort by the Association of Teacher Educators to assist teachers and administrators in what they plan and do about assessment. More rigorous scrutiny of teaching and teachers is essential, particularly to our membership, which includes those who prepare teachers, those who assess and evaluate teachers, and teachers themselves.

Assessment must be fair, appropriate to a context, carried on by qualified people, and consistent with

the larger issues in the functioning of good schools. My hope is that this monograph will help all who make and carry out policy achieve those goals.

Janet Towslee Collier
President
Association of Teacher Educators

Preface

Assessment (or evaluation) of teachers is a timely and controversial topic. Legislatures and state boards of education have instituted some type of teacher assessment in almost every state. The results, however, may not yield the hoped-for intent because too little thought has been given to the several purposes of assessment and the potential impact of various mandates.

There is no doubt that careful, thoughtful assessment can serve as one factor in the improvement of American schools. But it can also be a perfunctory exercise that neither applies high standards nor has a productive influence on the performance and the attitudes of teachers. Current efforts may also discriminate unfairly against minorities. Most important, a uniform system may not consider the particularities and the uniquenesses of the schools and the situations in which teachers find themselves.

The Commission on Teacher Assessment of the Association of Teacher Educators has written this monograph to explore many of the issues related to teacher assessment. The Commission hopes that the discussion will contribute to the deliberations of citizens, appointed and elected officials, and educators who have responsibility for thought, decisions, and action on this topic.

Theodore E. Andrews
Chair
Commission on Teacher Assessment

Introduction

Hilda, a college sophomore, has decided to become a teacher. To be admitted to a preservice teacher education program, she must pass a basic skills test in reading, writing, and mathematics. While she is in the program, she must complete student teaching successfully. To be recommended by her college for temporary certification, Hilda must pass a test of her knowledge of pedagogy and a test in her content area of specialization. When Hilda accepts a teaching job after graduation, she will enter an induction program. Regular certification as a teacher will depend on multiple performance appraisals conducted during her first year of teaching. As a second-year teacher, Hilda will begin a career-long appraisal process with periodic performance assessments focused on improving instruction, identifying areas for professional growth, assigning merit, and making employment decisions.

Teacher assessment (and evaluation[1]), as described above, is being proposed by various constituencies, seemingly as an all-purpose solution to many different problems:

- For *teachers*, assessment might raise teaching to a loftier status commanding more respect and admiration. With this attention might

[1]The Commission was charged with exploring the topic, teacher assessment. Members soon recognized that the concept of assessment covered only part of the appraisal of teacher performance being addressed in school districts and state legislatures. So the concept of evaluation was added. The reasons are discussed in detail on pages 32-40.

come greater rewards, which in turn would entice more-talented teacher candidates to a challenging profession.

- For the *general public*, teacher assessment at entry might serve as a screening process to ensure that only competent people were admitted to practice. Thus, only able teachers, more knowledgeable generally and professionally, and more skilled, would enter teaching. These successful teachers might bring about greater student learning.
- For *teacher educators*, assessment before admission to a teacher education program would reduce the need to accept and retain weak candidates and eliminate the possibility of boosting enrollment to receive more institutional support. Additionally, assessment at the completion of a program might be used to check the extent to which the teacher education curriculum had produced effective teachers, assuming that the test assessed what teachers needed to know and be able to do.
- Finally, for *local education agencies*, assessment might serve two purposes: to screen minimum competence and to identify areas for professional development. The results would give direction to staff selection and retention and could improve staff development programs, ultimately affecting student achievement.

Almost everyone agrees that some way should be found to ensure the competence of teachers in public schools. Yet little agreement exists on how to define a good teacher, how to carry out an assessment to ascertain the quality of a teacher, or

how to use assessment to nurture a teacher's growth throughout a career. This monograph describes and discusses a number of issues related to assessment that anyone concerned about teacher competence should consider.

There are three major phases in the continuous preparation of teachers, and assessment is appropriate in all of them. The first phase is the preservice period, before teaching. The second phase is the probationary period, usually three years long, during which a teacher is considered for a regular teaching license and tenure in a teaching position. (Most states grant a temporary license to graduates of approved teacher education programs; it must be converted into a regular certificate within a specified number of years.) The third phase is the career of the teacher, a period of perhaps 40 or more years.

Assessment serves different purposes in the first two phases than it does in the third. In both of the early periods, it is employed for screening and diagnostic purposes, providing answers to such questions as, What does this person need to learn to be a competent teacher? and Can this person demonstrate specified competencies at the performance level? In the third phase the emphasis is on providing feedback to the teacher so that he or she can improve professionally.

During the preservice period—that is, in the recruitment, selection, and preparation years—assessment criteria are applied to aid in finding, attracting, and preparing the best people available for teaching. The criteria help people interested in teaching learn what the job involves and help them and teacher educators discover if the required potential and commitment to teaching are present.

The criteria serve as goals for course work and field experiences and as standards against which accomplishments may be measured.

During the probationary period, assessment is used to help teachers identify and remedy strengths and weaknesses. It also serves to ascertain whether they can demonstrate minimum standards of performance.

Once a teacher is deemed sufficient—that is, once he or she has demonstrated minimum competence and been granted full employment status and a regular teaching license—assessment is used primarily to help him or her improve. The purpose and the best value of assessment then are to assist the teacher in fine-tuning skills and abilities and in staying in the forefront of new developments that contribute to teaching effectiveness—and thus to student learning.

Recently, assessment has been employed to determine eligibility for promotion in career-ladder and master-teacher plans. In some places, assessment is used to rate teachers for merit pay. More important, annual evaluations that may lead to dismissal are required in many states and school districts. The fact that relatively few teachers are placed on probation or dismissed is not as important as the tension the procedure creates within a school staff. Accountability measures in Texas and Arkansas, for example, have required experienced teachers to pass written basic-skills tests in order to maintain their positions. These measures have been criticized severely by teachers throughout the United States.

Such assessment practices have highlighted concern about how standards of effective practice

should be determined and how subsequent assessments should be conducted. Traditionally, standards and procedures for teacher assessment have been determined at the local and state levels. That practice has been criticized for two reasons: Practitioners have not been sufficiently involved, and procedures for enforcing standards have been inequitable, inadequate, and unevenly applied. A recent proposal to remedy these shortcomings of state and local teacher assessments is to create a national teacher-testing program. A position on this issue is not taken here, but some of the pros and the cons are given attention.

The Commission's Convictions About Teacher Assessment and Evaluation

Perhaps the best way for the Commission to begin a discourse on the what and the how of teacher assessment and evaluation is to state some of its convictions. They grow out of Commission members' collective experience and knowledge of the literature.

1. Paper-and-pencil measures have significant limitations. All such tests sample domains of knowledge, often in statistically random ways, making it easy to compare scores of individuals with those of the total group but not with an absolute standard. Also, researchers have yet to correlate a score on a written test with anyone's actual competence as a teacher.
2. Teaching behavior is complicated and intricate. Gathering an impression of the full scope of a teacher's skills and abilities may take months and requires more than two or three classroom visits.
3. Teaching involves a multitude of interactions and activities: several different instructional situations; curricular and cocurricular activities; working relationships with subordinates, colleagues, administrators, and other school employees; professional relationships with parents and other members of the com-

munity; participation in professional societies and associations; and other professional activities. A variety of data is necessary for a fair, objective assessment.

4. Assessment of teaching is a complex undertaking. It draws not only on information collected in classroom visits and paper-and-pencil tests, but also on data on a variety of activities and responsibilities. Paper-and-pencil tests best determine knowledge, not performance. Additional data are collected as teachers work with small groups and individual students, conduct teacher-student conferences, plan selection of materials, evaluate students' work and achievement, engage in curriculum development, and more. Interviews with teachers, inspections of their work plans and records, and observations of their effectiveness as members of a faculty also provide valuable information. In other words, "work samples" are a very appropriate source of data.

5. The context of a teaching assignment influences teaching performance. Observation may not give an accurate impression of a teacher's ability if certain factors in a school or a teaching situation make it difficult or impossible for a teacher to carry out an intended plan. For example, the lack of resources and materials, inadequate leadership and support, excessively disruptive students, personality clashes between teachers and students, or teaching outside their field can prevent teachers from achieving their potential in performance. Proper assignment is a continual challenge, one that

can be used to stimulate quality performance and provide greater variety in teaching.

6. Effective teacher assessment strives to be objective and devoid of bias and political favoritism. To ensure a somewhat objective and fair assessment, two or three evaluators should be employed and trained. They should visit a teacher several times during the year and complete a formal assessment. Teacher morale is boosted and evaluator objectivity is more probable when the assessment is conducted by professionals from buildings and districts other than those of the teacher being assessed.
7. Criteria included in the teacher assessment process are best understood and accepted when the people to be evaluated have a part in determining them.
8. Teachers take a variety of roles in the multifaceted job of teaching: instructor, tutor, examiner/evaluator, counselor, mentor, manager, employee, colleague, communicator to parent, member of a faculty and various school committees, individual professional, member of a profession, author of professional articles, and member of professional societies and unions. Other professional roles can include mentor to student teachers, member of a local or state curriculum committee, and member of a school or college accreditation team. All of the roles recognized in a particular district should be made public and considered for inclusion in an assessment.
9. Teaching, even though a public profession, is

a very personal one. Personality and performance are so integrated that they are essentially inseparable. Assessment outcomes become, at least in part, a judgment of personal adequacy. Therefore, the results of assessment should be documented and discussed privately with each teacher in a constructive way. They remain confidential, unless contested and required for evidence in a hearing. Hearings themselves should follow professional protocols of discretion and personal privacy.

10. Assessment of teaching, like assessment of most other human services, is subject to both error and bias. Observers must be trained to such a degree that error and bias are minimized. Also, there are different views on what constitutes good teaching. Therefore, teachers should have an opportunity to challenge an assessment that contains questionable criteria and anecdotal accounts. If necessary, there should be a hearing before peers and superordinates. Beyond local educational jurisdictions, professional practices boards are being established to ensure due process.

11. Schools and school districts vary in their expectations of teachers, and student bodies differ in socioeconomic level, cultural values, ethnicity, race, and other factors. Adequate teacher performance in one place does not ensure a similar performance everywhere. Therefore, assessment decisions and actions, inasmuch as they relate to employment and improvement, should be kept separate from decisions and actions about licensure.

Issues in Assessment

A fundamental measurement problem related to assessment is determining what to assess. What does a teacher need to know and demonstrate in the classroom to be considered competent?

After it has been determined what abilities, knowledge, skills, and attitudes a teacher should possess, there is the problem of deciding how to collect the data and where to set the passing level. Will competence be determined by performance in the classroom or more simplistic measures of accountability (e.g., grade-point averages or scores on standardized tests)?

If the teacher must demonstrate competence in reading and writing, what will be considered the minimum standards of performance? Will they be the same for all teaching levels and teaching areas? Who or what group will set the standards? Will there be state or national standards? Will there be provisions for remediation? If a person passes one portion of a test and fails another, will he or she repeat only the failed portion?

Further problems concern deciding who will be responsible for testing, keeping records, etc. Will the teacher candidate bear the cost of test administration and scoring? Who will support the administration and personnel to carry out the program? Will local or regional teams or existing state department and university personnel assume these responsibilities? Once the question of adequacy is satisfied, how will assessment be used throughout

a teacher's career to foster continuous professional development?

These are not easy questions to consider, and they have no simple answers. Currently, school and state department personnel, state legislators, educational researchers, measurement specialists, and the courts are being asked to furnish the answers. For the most part, the actual decisions are being made by state policy makers (legislators or state education department personnel), frequently on the basis of what is the most politically feasible. In the following pages the Commission examines some major research, measurement, political, and legal issues related to assessment. It also discusses the often-neglected issue of teacher evaluation.

Research Issues

A critical problem when assessment is carried out to support administrative decision making (e.g., in relation to hiring, firing, and promoting) is substantiating the validity of the competencies included in the assessment. Researchers over the past 20 years have completed a significant number of studies identifying specific teaching behaviors that are related to student achievement. Educational policy makers often wish to incorporate these findings directly into assessment policies. Where the data are consistent and the findings comprehensive, this is educationally sound. However, the amount of information that researchers have validated about effective teaching is insufficient compared with the amount that still needs to be learned. In addition, the following caveats pertain to the use of research studies:

1. Research studies vary in a number of significant ways (e.g., the number of teachers in the

study, the grade level, and the socioeconomic status of students), even though they may all be directed toward the same goal—in this case, attempting to discern what the teacher does that contributes to increasing the learning and the satisfaction that occur in the classroom. *Generalization of research findings to groups and situations beyond those in a particular study should be cautiously considered.*

2. *The interpretations placed on teacher-effectiveness research findings should be carefully studied.* Clearly they can serve as a source of guidance to teachers or administrators, to use in identifying and interpreting classroom events. Often, however, a generalization is inappropriately applied to all students. Effective teachers may need to use different behaviors with low-achieving students than they use with high-achieving students. For example, in a study reviewed by Medley (1982), the patterns of teacher behavior that were effective with high-ability students in reading were ineffective with low-ability students in reading.

3. School districts are using teacher-effectiveness research to develop observation procedures or instruments in an attempt either to improve the performance of a teacher (formative evaluation) or to determine if a teacher will be certified, hired, fired, reassigned, selected for merit pay, or promoted (summative evaluation). *It is essential that the validity and the reliability of the observation/evaluation instruments used for such purposes be clearly established and that the observations be conducted by a trained and reliable observer/evaluator.*

4. Some of the theoretical issues in teacher assessment concern the basic assumptions that underlie the research used to examine teaching. Fenstermacher (1978), for example, criticizes most of the research on teaching on the grounds that it "lacks a normative theory of education." It does not, he contends, "incorporate a conception or justification of what is ultimately worth knowing and doing; it includes no view or defense of right conduct or moral integrity, nor does it give consideration to or argument for the ethical obligations and reasonable expectations of persons who act in specific, historical, social, political, and economic settings" (p. 175). Winne and Marx (1977) argue that research on teaching "has been devoid of postulating a reasonable psychological foundation mapping out why certain events of teaching should influence particular learning processes in students which are manifested later in observable behavior such as test responses" (p. 670). In short, teacher-effectiveness research, according to some critics, has not addressed value questions nor dealt with the theoretical assumptions of why a particular style of teaching influences student learning. *Developers and users of teacher assessment systems should be aware of and consider these theoretical and value questions.*

5. Popkewitz, Tabachnick, and Zeichner (1979) direct their attention to another important criticism of teacher-effectiveness research that has parallels in teacher assessment. In their view the research "has certain limitations for educational inquiry and for understanding the phenomena of teacher education" (p. 52).

> They argue that *observational approaches such as those used in teacher assessment take a conservative orientation to the tasks of professional preparation and distort the phenomena under study.* Furthermore, they maintain, the approach has a behaviorist orientation that assumes "it is possible to identify, control and manipulate specific outcomes of teaching by altering selected aspects of a teacher's overt behavior" (p. 52).

Any profession seeking to improve practice and the assessment of it will produce some knowledge. That knowledge will comprise both hard and soft theories; that is, some will stand any test to which they may be put, whereas others will be tentative and caution must be applied in their use. People involved in teacher assessment must, first, know the knowledge base and, second, be able to discriminate between what is known for certain and what is thought to be known. Some important considerations in reviewing research are as follows:

- The results of a study are not generalizable beyond the population from which the study sample was drawn.
- Results obtained from correlational studies do not indicate cause-and-effect relationships between variables (e.g., teaching behavior and student achievement).
- More of something is not always better—some relationships are not linear, but curvilinear.
- The assessment instrument used in a study must be valid and reliable, and the classroom observer must be reliable in his or her observation.

- Studies of children in one subject, at one grade level, from one socioeconomic level, may not be generalizable to other children.
- Results from one study are often not replicated in other studies.
- A margin for error is always present in an instrument used to collect data.
- The levels of significance set by researchers vary considerably.
- The number of students and teachers included in a study is important in understanding the extent to which the findings are generalizable.
- Much research completed in the area of teacher effectiveness has been correlational in nature.
- Teacher-effectiveness research usually ignores theoretical or value questions.
- The "whole" of teacher education may not be assessable "part" by "part."

The above caveats are presented as a caution against overgeneralizing research findings. A number of these caveats are addressed in more detail in the next section, Measurement Issues.

Measurement Issues

For any assessment instrument such as a written test or an observation system to be educationally or legally defensible, it must be valid. That is, the instrument must actually measure the quality/qualities for which it was designed. In addition, the instrument must have predictive validity; that is, it must provide assurance that teachers who effec-

tively demonstrate the behaviors on the instrument can indeed perform those behaviors. Many observation instruments are being designed on the basis of research findings in order to provide predictive validity. The Florida Performance Measurement System (FPMS) is such an instrument. Preliminary findings from a major study of practicing teachers indicate that the FPMS may well have predictive validity (Peterson, 1986). Achieving even a moderate degree of predictive validity may be very difficult because many factors have to be controlled. It is so difficult that the courts, in adjudicating disputes, have consistently avoided ascertaining predictive validity in favor of verifying content validity.

In the absense of research to support the inclusion of certain teaching characteristics on an observation scale, a systematic procedure should be used to establish the content validity of the characteristics. Content validity is founded on the best judgment of teachers, administrators, teacher educators, and, in some cases, the courts. Although some state systems are being carefully built, local districts seldom use a rigorous process. In most urban school districts, the local teachers' union and the school system administration negotiate the content and the procedure of the assessment process.

An assessment procedure should be job relevant. Instruments employed for the assessment of teacher competence should reflect the types of knowledge and skill required to perform satisfactorily in a particular position. When a teachers' union and a school administration negotiate a teacher assessment contract, there is a concerted effort, particularly on the union's part, to be sure the characteristics that become part of the observa-

tion system relate directly to what teachers do in the classroom. Although all the sacred canons of content validity may not be observed, the agreed-on assessment process usually reflects a low-to-moderate level of content validity. That is, the assessment process derived in this manner may be more job relevant than one that is not negotiated. Content validity and job relevance are serious issues that cannot be ignored, as affirmed by court cases cited later in this monograph.

In addition to being valid, the instruments used to collect the data must be reliable. That is, there must be consistency in the results reported by different observers or by the same observer on different occasions. Reliability of a teacher assessment process assumes that the teacher will receive the same score each time the process is applied (e.g., each time he or she is observed). Researchers have found that the inconsistency of the relationship between teacher behaviors and student outcomes often can be attributed to the unreliability of the observers (raters) of classroom behavior and the infrequency of observation. Several well-trained raters and a number of occasions for observation are required to measure certain variables reliably. This important issue is often not adequately addressed in teacher assessment plans.

A major issue raised by Stodolsky (1984) is the assumed stability and consistency of teaching. Too many observers and evaluators of teaching generally perceive a teacher as effective or ineffective, and believe, Stodolsky states, that effective teaching is "characterized by one set of behaviors that are context free" (p. 11). Stodolsky's research clearly demonstrates that systematic variation in teaching and instructional arrangements is com-

mon for many teachers. This disclosure precludes the application of one set of rater characteristics in the evaluation of teacher effectiveness. Several observations of a teacher in different instructional situations are necessary. Meetings with the teacher following the observations may be much more productive if the evaluator understands that instruction might vary considerably in different situations.

Another important issue mentioned earlier is the establishment of criteria of effective performance. If, for example, clarity, enthusiasm, questioning techniques, and monitoring of student work are some of the significant characteristics of teaching, an evaluator might be asked to determine what level of performance will be considered satisfactory. What type of scale(s) will be attached to these characteristics? Can a teacher fall short on 4 of 10 characteristics, yet receive an overall satisfactory rating? Are some characteristics more critical than others? The development of criteria of acceptable and unacceptable performance is an important measurement problem. A major reason for focusing on formative observation instruments is that passing scores do not need to be established; preservice and inservice teachers receive feedback and work toward improvement in a nonthreatening environment.

Evaluators and administrators often fail to consider the sometimes complex interrelationships between the extent of use of specific teacher behaviors and student outcomes. Teacher praise, as an example, has been found to be an effective teacher behavior, but too much praise has been associated with deleterious effects (Brophy & Good, 1986). In most situations, more of a certain behavior is not

better. This form of relationship is important to understand because many behaviors may have a point of diminishing positive return.

As noted above, effective teaching may be specific to a context. Similarly, particular teaching behaviors affect different types of students differently. Teacher behaviors are often viewed independently of the curricular context in which they occur, and little attention is paid to the continuing nature of classroom activities that also affect an assessment procedure. Routines introduced by an effective teacher (e.g., rules for classroom behaviors and procedures that maintain instructional momentum) may be so well established by mid-year that they are not readily observable and not amenable to assessment.

Political Issues

Testing and performance-appraisal issues are directed at one of the core political questions in education: Who controls the teaching profession? Although the authority to recommend candidates is often delegated to teacher education institutions, states have long controlled entry into the profession through the certification process. In some states the maintenance of professional standards throughout a professional's career is also affected by certification standards, which are increasingly becoming tied to assessment requirements.

Trends toward increased assessment at all stages of the teacher's career, whether through paper-and-pencil tests, performance appraisal, or some other means, have shifted control of the profession downward and outward. For example, the current trend to require performance assessment of begin-

ning teachers before standard certification shifts authority to recommend candidates for certification from teacher education institutions to local school districts. This disperses the responsibility to a larger number of agents and essentially diversifies teacher education programs leading to certification because local on-the-job experience will affect recommendations.

A second key political question is, Who should control the allocation of resources for education? When legislators or state policy makers decide that funds will be designated for the development and the administration of tests and performance appraisals, resources available for other educational needs are often reduced. In addition, state-mandated induction and alternative-certification programs have special implications for the use of resources at the local district level because they place both increased power and increased responsibility in the local district. The increased responsibility requires greater resources at the local level for supervision and evaluation of beginning teachers and candidates for alternative certification. Those resources may or may not be forthcoming from the state; reallocation of finite local funds or tax increases may be required.

Research on teaching, the mission of schooling, and a philosophy of education, *not* the assessment process, should drive and direct how teachers are prepared, as well as how they teach in their classrooms. The danger is very real: If the specifications for an exit exam emphasize direct instruction to the exclusion of other methods, teacher preparation will eventually reflect the same emphasis because the payback resides there. Indeed, preparation programs that do not provide candidates with an

opportunity to learn what is required in the state's assessment system risk being labeled irrelevant. Likewise, if the classroom appraisal process focuses only on direct instruction, the teacher may use other methods of instruction less often, even when they might be more appropriate. When the teacher's impact on cognitive achievement is the sole learning being measured by the assessment system, then social learnings, values, and affective development will probably be ignored.

Impact on minorities. Perhaps the most critical political issue is the impact of state assessment programs on minorities. Following are some illustrations:

- In 1986, 64.9 percent of the Native Americans and 60.4 percent of the blacks who took the Arizona Teacher Proficiency Examination failed it on their first try, as did 39.3 percent of the Hispanics. By contrast, 26.1 percent of the Asians and 18.2 percent of the whites failed (Ingrassi, 1987).
- South Carolina has developed a basic skills test called the Education Entrance Examination to assess students seeking admission to teacher education programs and beginning trades and industries teachers seeking continuing contracts. Cumulative results of eight administrations of the test (fall 1984 through summer 1987) have revealed that 77.1 percent of the 1,040 black students who have taken the test have failed on their first try, whereas 22.2 percent of the 6,354 whites have failed (South Carolina Department of Education, 1987).
- Since 1980, Florida has administered a teacher certification examination, which all students

must pass in order to receive an initial teaching certificate. In recent years, failure rates for minorities, on their first attempt, have averaged about 60 percent, compared with about 10 percent for whites (K. Loewe, personal communication, December 14, 1987).

- A five-year summary of Georgia's teacher testing program revealed passing rates of 34 percent among blacks, 87 percent among whites ("Few Blacks Passing," 1983).
- Among those who took the first California Basic Education Skills Test in 1983, passing rates were 26 percent of the blacks, 39 percent of the Mexican Americans, and 76 percent of the whites (Gifford, 1987). Figures for 1985 showed some improvement: 33 percent, 46 percent, and 81 percent, respectively (Connell, 1987).

At the same time, the number of minority students in America's public schools is significant, and is rising rapidly in many states. In the state of Washington, the overall population increased by approximately 20 percent between 1970 and 1980. During the same period, the Hispanic population grew by 37 percent, the black population by 30 percent, the Native American population by 67 percent, and the Asian population by 93 percent (SPI, 1984). Competency tests may significantly reduce the number of minority teachers.

Several national programs to provide remediation to minority college students have been developed, including one at the American Association of Colleges for Teacher Education. There is some evidence that, with proper coaching and resources, minority students can improve their performance

Table 1. Percentages of Graduates of Historically White and Black Virginia Institutions Scoring Below the Derived and Adjusted Standards for Selected Tests

			Derived Standard		Standard Minus 1SE		Standard Minus 2SE	
Examination	*Inst. Type*	*Sample Size*	*Scaled Score*	*% Below*	*Scaled Score*	*% Below*	*Scaled Score*	*% Below*
Core Battery								
Communication Skills	White	1,661	651	9	647	4	644	3
	Black	89	651	58	647	49	644	44
General Knowledge	White	1,661	641	4	637	1	634	1
	Black	89	641	48	637	40	634	32
Professional Knowledge	White	1,661	641	3	637	1	633	1
	Black	89	641	39	637	30	633	17
Area Examinations								
Elementary Ed.	White	443	538	14	512	8	485	3
	Black	40	538	54	512	39	485	24
Early Childhood Ed.	White	816	515	11	485	6	455	3
	Black	59	515	65	485	54	455	39
Industrial Arts Ed.	White	54	599	40	579	33	561	3
	Black	19	599	82	579	73	561	57
Physical Ed.	White	329	589	30	561	20	532	13
	Black	37	589	85	561	73	532	54
Business Ed.	White	64	575	22	547	14	519	6
	Black	17	575	74	547	49	519	29
Ed. Mentally Retarded	White	320	549	27	521	17	492	11
	Black	59	549	70	521	57	492	43

Note: From *Validation Study of the National Teacher Examinations for Certification of Entry-Level Teachers in the Commonwealth of Virginia* (p. 19) by L. H. Cross, 1984, April, paper presented at the annual meeting of the American Educational Research Association, New Orleans. Reprinted by permission.

dramatically (e.g., Cross, 1987c). The problem appears to be related more to black students who attend black institutions than to those who attend racially integrated or predominantly white institutions (Ayres, 1983; Cross, 1987a, 1987b).

A concern for the impact on minorities also has a major influence on the cut-off scores established to distinguish between those who pass and those who fail a competency test. Elaborate procedures that involve review of individual test items (e.g., the Angoff, Ebel, or Nedelsky methods, described in Livingston & Zieky, 1982, pp. 15-31) are used to establish cut-off scores, which are then often lowered to lessen the impact on minorities. After Cross (1984) completed a validation study of the National Teacher Examination for the Commonwealth of Virginia, he recommended that the cut-off scores be lowered (i.e., that they be placed two standard errors below the derived standards). He reported, "Graduates from historically black Virginia institutions will be placed at considerably greater risk in comparison to the graduates of historically white institutions, regardless of where the cut-off scores are set" (p. 18). This study, the results of which appear in Table 1, indicated that if the cut-off scores were not lowered, 3-9 percent of students from white institutions would fail the Core Battery, compared with 39-58 percent of black students, whereas if the scores were lowered, 1-3 percent of students from white institutions would fail, compared with 17-44 percent of students from black institutions. In practice, the major result of the test may be to reduce the number of minority (black) teachers while having no impact on the number of students entering the profession from predominantly white institutions. Norfolk State College

Dean Elaine Witty, testifying before the Virginia State Board of Education, stated, "If the goal is to lower the number of black teachers, then use of the National Teacher Examination or any other achievement test would be satisfactory" ("Educator Implies," 1984, p. C-1).

There are no simple solutions to alleviating the impact that state testing programs are having on minorities. But recognizing the problem is clearly the first step.

Legal Issues

America has been characterized as an adversarial society, and occupational assessments, primarily written tests for the purpose of licensing, hiring, or other personnel decisions, have been the objects of considerable litigation. Up through the early 1970s, tests for prospective or employed teachers were routinely struck down because of discriminatory impact or, in several instances, discriminatory intent. The late 1970s, with key cases in North and South Carolina, saw a shift in judicial interpretation. The court in *U.S. v. South Carolina* (1977, 1978) upheld the practice of teacher certification testing when supported by evidence of content validity and judgmentally determined passing scores, despite disproportionate impact on minority candidates. The decision was summarily affirmed by the U.S. Supreme Court.

Each state has long had the authority to regulate, by examination or otherwise, entry into professional practice (*Dent v. West Virginia*, 1889) and since 1975 has had the authority to regulate entry into teaching specifically (*U.S. v. North Carolina*, 1975, 1977). States also have the right to take steps

to maintain professional competence, as long as they do not violate the constitutional rights of individual practitioners (*Scheelhaase v. Woodbury Central Community School District*, 1972, 1973, 1974).

Most litigation has been brought under the 5th (due process) and 14th (due process and equal protection) Amendments to the U.S. Constitution, the Civil Rights Act of 1964 (Title VII, which prohibits discrimination), and most recently the *Uniform Guidelines on Employee Selection Procedures* set forth in final version in 1978 by the federal Equal Employment Opportunity Commission (referred to henceforth as the EEOC *Guidelines*).

What follows is a set of 10 principles that policy makers deliberating on teacher assessment programs might usefully bear in mind (along with appropriate psychometric and validation principles). The principles were derived from court decisions of the 1970s and 1980s.

1. *The "rational relationship" standard.* According to this principle, the content of a test for teachers must have a logical relationship to teaching effectiveness (a legitimate state interest). Tests of subject-matter specialty or of basic reading, writing, and arithmetic skills would have such a rational relationship; measures of artistic or athletic ability applied to all teachers generally would not. A relatively easy standard to meet, it has been applied to examinations for a number of different vocational and professional occupations, and they have typically been upheld (e.g., *Tyler v. Vickery*, 1975, 1976; *U.S. v. North Carolina*, 1975, 1977; *Washington v. Davis*, 1976; and *U.S. v. South Carolina*, 1977, 1978).

2. *Job-relatedness of the examination.* Beginning with the 1971 *Griggs v. Duke Power* case, when employment decisions (e.g., hiring) disproportionately exclude minorities (and thus have so-called adverse impact), employers must demonstrate the job-relatedness of the selection criteria. This requirement is reflected in the EEOC *Guidelines* and was applied in the *Guardians Association v. New York Civil Service Commission* case (1980, 1981). Without doubt, a key requirement now is that teacher assessment instruments be validated in terms of job-relatedness or job relevance.

3. *Suitable test validation.* Evidence of job-relatedness or of other relationships between a given test and teacher training and/or performance criteria is typically produced by means of a validity study. A line of court decisions, including the North and South Carolina cases and *Debra P. v. Turlington* (1979, 1981), has upheld assessment programs that have been validated to the courts' satisfaction.

   Evidence of a variety of types of validity (discussed in an earlier section) is acceptable, for example, under the EEOC *Guidelines*. Evidence of content validity of teacher tests, established through judgments of item content by panels of educators, was viewed as appropriate by the court in the South Carolina case. Technical care in conducting the validity study and good documentation of the steps involved have also been regarded as important (*Albemarle Paper Co. v. Moody*, 1975). Most states using tests for

teacher certification have validated their instruments in ways sanctioned by the South Carolina decision.

4. *Nonarbitrary passing scores.* The process used to establish a passing or cut-off score is closely related to a test's content validity. Virtually all states with teacher-certification assessment programs have established such cut-off scores for their respective states. What is critical from a legal standpoint (e.g., the North Carolina case and *U.S. v. Nansemond County School Board*, 1972, 1974) is that the passing score not be set in an essentially arbitrary manner. For example, a board or a commission cannot legally adopt a passing score of 70 percent in the absence of any other information.

   Typically, data for setting passing scores are gathered at the same time as content-validity data are gathered. A number of different procedures (named for their developers—e.g., Angoff, Ebel, and Nedelsky) are acceptable; all involve ratings by carefully selected panels of educators (see Livingston & Kieky, 1982, pp. 15-31). Such empirical methods of determining cut-off scores have been consistently upheld by the courts beginning with the North and South Carolina decisions.

5. *Absence of discriminatory intent.* The Constitution, the Civil Rights Act, and considerable case law combine to prohibit intent to discriminate by race or sex in occupational assessment. Discriminatory intent, however, cannot be inferred from disproportionate or "adverse" impact (in the language of the EEOC *Guidelines*). For example, the

fact of different passing rates for different racial groups on a state's teacher certification examination does not necessarily mean that the state intended smaller proportions of particular racial groups to pass. In the South Carolina case the "rational relationship" rule was applied (together with no evidence of discriminatory intent), and the assessment program was upheld, in spite of lower passing rates for black candidates.

6. *Disproportionate impact and the need for suitable evidence of validity.* As discussed elsewhere in this report, on most tests of knowledge or cognitive ability, minority teacher candidates have scored somewhat lower than whites, on the average. This is indicative of disproportionate or "adverse" impact. (Adverse impact exists, according to the EEOC's rule-of-thumb, when a particular group's passing rate is less than four-fifths of the highest-performing group's passing rate.) Whether or not a state or an institution has reason to believe that there will be disparate impact from a certification or admissions examination it proposes to adopt, the instrument must be carefully validated for job relevance (*Griggs v. Duke Power*, 1971; *Albemarle Paper Co. v. Moody*, 1975; *Connecticut v. Teal*, 1982; and the Guardians case, 1980, 1981, which was brought under the EEOC *Guidelines*).

7. *Technical soundness of assessment instruments.* The court in the Guardians case wrote that "the test makers must have shown reasonable competence, thoroughness and care in constructing the test itself." With regard to test reliability, the court indicated that

"without some substantial demonstration of reliability, it is wholly unwarranted to make decisions . . . that turn on one-point distinctions . . ."

8. *Inadvisability of rank-ordering on the basis of test scores.* Some jurisdictions, particularly school districts, have followed the practice of ranking applicants from highest to lowest test score and then making hiring or other personnel decisions on this basis, as opposed to forming a pool of eligible candidates whose test scores have met a preset minimum passing score. The court in the Guardians case cautioned against this practice in the absence of acceptable evidence of test validity.

   Logically, there needs to be evidence of criterion-related validity rather than content validity. That is, data are needed to show that very high test scores are associated with outstanding teaching performance, somewhat lower scores with somewhat less than outstanding performance, and so forth. Several decades of teacher assessment research do not hold promise that such strong correlations can be obtained.

9. *Inadvisability of using a test as the sole decision criterion.* Present-day mental measurement is far from perfect. Assessment instruments are not perfectly reliable in the sense that a candidate will obtain identical scores each time the test is taken. Nor are they entirely valid in the sense of a one-to-one relationship between test score and on-the-job performance. In the case of *U.S. v. Nansemond County School Board* (1972, 1974) the court

ruled that teachers cannot be discharged solely on the basis of a test score.

10. *Adequate notice to candidates.* Teacher candidates faced with tests or other assessments presumably have a due-process right to advance notice, as has been accorded students facing competency tests (*Debra P. v. Turlington,* 1979, 1981). No rule-of-thumb has emerged; lawyers consulted on the matter advised that two years' notice is probably appropriate. It is also advisable, based on *Debra P. v. Turlington,* to provide candidates with opportunities to retake an examination if they initially fail it, and to obtain remedial assistance.

An Often-Neglected Issue: Evaluation to Improve Teaching

Although evaluation[2] designed to improve teaching[3] is mentioned early in this volume, most of the discussion implies standardized teacher assessment methods and techniques. The current preoccupation in the country with standardized teacher assessment and the concern that such appraisal be useful, fair, and objective are the reasons for that emphasis. However, the importance of evaluation designed strictly to improve teaching should not be minimized. Such evaluation is essential to school improvement.

---

[2]The term *evaluation* is used here because it better fits the appraisal-and-review process discussed.

[3]The term *teaching* is broader than the term *instruction.* It includes all the activities teachers undertake as a part of their regular employment.

Evaluation to improve teaching is clearly different from standardized assessment. Assessment is measurement of competence. Evaluation is continuous appraisal, reflection, analysis, rethinking, improving, and refining of teaching practice. Teachers participate in evaluation alone and with colleagues, supervisors, mentors, and administrators. The word *evaluation* is not adequate for the scope of the concept. Yet no other term encompasses the several processes that are a part of deliberate growth in teaching.

Standardized assessment can provide quality safeguards in preservice teacher education, by determining who enters and becomes fully certified in the profession and who becomes tenured. Evaluation can ensure the maintenance of a vital, dynamic teaching force and provide tremendous personal and professional satisfaction to the teacher.

Teaching is a human service, different in both purpose and method from work in industry, business, or government. It can use or adapt some of the approaches and processes of appraisal employed in other occupations, but in many ways it has distinct characteristics and requires unique procedures and methods of appraisal. Teaching is a very imprecise endeavor. Few if any standard methods or techniques ensure learning for all students in all situations. Teaching and learning do not have a direct cause-and-effect relationship. Learning evolves over time. It is often a matter of degree. Teaching influences and affects learning and depends on the teacher's diagnosis of students' motivation, ability, and attitude to determine what might be learned and how. Context, often outside the control of the teacher, is a major

influence on the teacher's ability to teach and the student's proclivity to learn. Teaching is often directed at groups, yet individual differences in students (which are wide and deep) require continuous attention and differentiation. The range of achievement among students in the same classroom is usually great. Teachers must recognize, ascertain, appraise, and reward each student's achievement in appropriate ways.

Teaching includes more than instruction before a class. When students are involved in activities such as small group discussions, critiques of each other's work, and hands-on projects or experiments, the teacher may be in a supportive, facilitating role, serving as counselor and resource person. In such instances, observing teaching practice is difficult or impossible. The teacher also has several other roles: member of a faculty, colleague to fellow teachers, manager and guide for teacher aides and student teachers, association and union member, and member of a profession (California Commission, 1964; Edelfelt, 1979).

Leaders in school districts throughout the country work with teachers every day to help them remain vital, to assist them in improving school program, and to support and encourage their career-long professional development. Such work is complex and requires a keen awareness of the needs of teachers. The quality of the interaction between supervisors and teachers varies greatly. Skill and sensitivity in working with teachers are not as widespread as one might wish; supporting teachers in examining their teaching and helping them translate evaluation and analysis into better teaching is exceedingly difficult. Almost all school administrators, however, are greatly interested in

finding more effective ways to work with teachers to improve teaching. Attention to this type of evaluation is critical because it is an ever-present part of school improvement.

Evaluation in the best of settings. Under ideal conditions of practice, evaluation involves supervisors, principals, mentors, advisors, and (sometimes) peers observing and reacting to teaching. [A recent study (Bacharach, Bauer, & Shedd, 1986) commissioned by the National Education Association found that teachers claimed to learn more from peers than from other sources.] Moreover, evaluation is seldom an isolated activity. Rather, it is part of a larger, holistic process, addressed in conjunction with many other factors that influence teacher behavior: curriculum content, instructional method, the abilities and interests of students, the personality of the teacher, the group dynamics of a class, the social system of the school, physical and psychological environment, and, of course, the effectiveness of the teacher. The observer does not make judgments using one set of criteria that applies to all teachers. Instead, criteria for the individual teacher are developed based on school goals, the teacher's intent, the capacities and the interests of students, and the learning environment.

The observer (supervisor or peer) is a listener, a helper, a guide, a mirror for behavior, a mentor, an advisor, a coach, and a counselor. A high degree of trust exists between the teacher and the observer. The teacher frequently invites the observer to view an event, to sense a climate, to observe interaction between teacher and students or among students themselves. The teacher often seeks opinions on such questions as, How well is this activity work-

ing? Is the environment right? Am I using the best materials? Is the content appropriate? Or the teacher may want to know, Am I reaching all the students? Should different methods be used? Am I too directive? Do students have adequate encouragement to think for themselves? On occasion (frequently in difficult schools), teachers want to know why a student does not seem to be interested or is misbehaving. Teachers also worry about whether the rhythm and the timing of the teaching-learning process are appropriate.

The observer sometimes reacts only to the questions the teacher raises. Even when a teacher asks, What else might be improved?, the observer is perceptive enough to sense what the teacher is able to hear or discover through reflection. The observer listens carefully and relies heavily on the teacher's introspection, on his or her reflection on questions that the interaction has triggered.

The observer encourages self-evaluation. All of the above questions can foster self-assessment. But it can be triggered in a number of other ways. For one, there is recorded feedback. Where videotaping or audio taping is available, teachers can be encouraged to look at and/or listen to themselves. If a teacher feels comfortable viewing tapes with an observer, two perceptions can be brought to bear on analysis and evaluation.

Another method of self-evaluation is keeping a journal (Holly, 1984). Reflective writing requires thought and discretion to report objectively and candidly how one feels about, how one perceives, what has happened in the classroom. Reviewing a description or a reaction to an event causes a teacher to evaluate his or her perception of the situation.

Documentation, closely allied to journal-keeping, is another technique for evaluation (Carini, 1975; Engel, 1975). This procedure calls for more than written reflection by the teacher. It includes collecting other data and information, such as vignettes and portraits of classroom and school activities, specimens of student work, student reactions, photographs—anything that helps record the events of teaching and learning.

Written data and other documentation over time provide a record that unchronicled experience can never match. Learning takes time. It cannot be seen in a few classroom observations spread over a year. Long-term behavior demonstrates more accurately what has been learned. Such data provide longitudinal perspective. A journal and documentation give that kind of scope. When a teacher is willing to share a journal or other records with an observer, the discussion of teaching can be greatly enriched.

This kind of record-keeping also provides vivid data to share with students and parents. It photographs, symbolically if not actually, aspects of the teaching-learning process in school. It makes vignettes and slices of classroom life, as well as descriptions of what is being taught and learned, available for discussion and evaluation.

In still another technique, *triangulation*, the perceptions of a teacher, his or her students, and a professional observer are collected, compared, analyzed, evaluated, and interpreted by the teacher and an advisor. Some insights possible for the teacher in such a process are reported by Elliott (1977).

The art of teaching. Attention to artistry and intuition in teaching, to the exercise of almost mysteri-

ous talents, is also a part of evaluation for the improvement of teaching. Why a teacher lingers on a particular question, opts not to use a prepared lecture, changes the pace of a lesson midstream, decides to have a particular student read her essay aloud, moves to small group discussion, alters an assignment, instructs the whole group, is not always explainable. Looking back, the teacher typically reveals, "I felt that was the right thing to do." A rationale can usually be developed after the fact, but the teacher's moves amidst hundreds of cues that he or she processes while teaching are pure artistry. The teacher orchestrates learned techniques and adds intangibles, and the performance is an art.

Recognizing the art of teaching, analyzing the chemistry of the situation, discussing what was viewed, and evaluating it require artistry on the part of a supervisor or a peer. The usual systematic analysis, examination of a plan, reaction to its execution, and rational discussion and feedback are often insufficient. Discussing how one feels about an artistic performance is usually inadequate. It is only part of the story. Acknowledging body language and facial expression, reviewing oral descriptions of what went on, and admitting that one does not know why a teaching episode evolved as it did, all may help. But articulating reaction may be as difficult as analyzing a Monet painting, an Itzhak Perlman performance, or a Dylan Thomas poetry reading.

The outcome of this kind of supervision and evaluation is not a numerical rating or a judgment that something was correct and something else wrong; the results are never that simplistic. They are more significant than standardized assessment

in influencing teacher behavior for several reasons. First, less threat is involved. The teacher is not being graded. The assumption has been made that the teacher is competent. The intent is to gain insight into the interaction of teacher and students, to analyze and refine teaching skills, to improve the many conditions that influence the participants, and to examine the context in which teaching and learning take place.

Second, the teacher knows the observer and his or her persuasions on teaching. There is mutual respect. The result of observation and discussion is understanding and insight, not a verdict or correction; it is the enhancement and development of professional judgment. The current emphasis on experienced teachers serving as mentors to beginners illustrates well this kind of a relationship (Galvez-Hjornevik, 1986).

Third, the outcome is not advice on what should be done differently. It is help in recognizing how well some approaches and activities are working and where improvement is needed. Both the teacher and the supervisor realize that what was successful last year or in another class may not be appropriate for a current class. *There are no set ways to teach that work all the time.*

Fourth, professional confidentiality prevails. There is no public announcement that a teacher did or did not pass muster, get promoted, or receive a particular rating.

Fifth, evaluation covers much more than performance before a class, a lesson being taught, the dispensing of information, didactic teaching (this in contrast to standardized assessment, which spotlights the teacher's behavior before a class).

Evaluating is a broader concept than *assessing:*

- Evaluate: "to determine the significance or worth of usually by careful appraisal and study" (*Webster's Ninth,* 1986, p. 429)
- Assess: "to determine the importance, size, or value of" (*Webster's Ninth,* 1986, p. 109).

Evaluation is a holistic endeavor, encompassing the engagement of students—their observable interests, attitudes, attention span, motivations, interactions, industry, and inquisitiveness.

Recommendations

In concluding this review of issues related to teacher assessment, the Commission offers five basic recommendations:

1. Teacher assessment should be an integral part of the larger enterprise of education. That is, it should reflect the values and the purposes of the school and be consistent with the philosophy of the school, with notions about how learning takes place, and with the process and the expected outcomes of education.
2. A program for teacher assessment should have at least five purposes:
   a. To offer clear guidelines for the development and monitoring of preservice teacher education programs
   b. To screen candidates for regular licenses
   c. To serve as a diagnostic device for inservice professional development
   d. To gather data in cases of alleged incompetence and/or malpractice
   e. To provide information for the public.
3. Philosophies and assumptions that guide item selection for teacher tests and teacher performance appraisals should be articulated for the public and for the profession.
4. The assessment method chosen should be appropriate for the goal and the objective

being measured. If public confidence in teachers' reading, writing, and arithmetic skills is lacking, a criterion-referenced paper-and-pencil-test of those skills will indicate whether the concern is well-founded or not. The same test will not accurately predict classroom teaching skills. In other words, a paper-and-pencil test will not tell the public whether or not a teacher is a good teacher.

5. Performance-appraisal systems should be carefully developed. The public must be made aware that assessment and performance appraisal of teachers are not easy solutions to complex problems in education, but are, in fact, part of a complex solution to complex problems. Personnel with expertise in teaching, testing, and appraisal must diligently and deliberately develop, implement, and revise these processes. Such an evolutionary endeavor will require both time and money.

The teaching profession is at a difficult time in the evolution of productive teacher assessment and evaluation. People want these processes to have objectivity, reliability, validity, and rigor. Yet they want teachers to be open, creative, original, insightful, sensitive, and humane. Assessment and evaluation that are characterized by the former are defensible and essential, but they must not preempt teachers' becoming the latter.

If assessment and evaluation are to promote both of the above (the Commission suggests that is possible and desirable), assessment to ascertain degree of competence needs to be separated from evaluation to help teachers stay vital and dynamic. Teaching is so taxing a career that nurturance may

be more important to teachers than to most other professionals.

The emphasis, however, must become more than assessment or evaluation of teachers per se. Consideration of the context in which teaching occurs—the social system of the particular school or school system in which a teacher teaches—is absolutely essential. If both the profession and the public were to identify the critical factors essential to a good school and assess (and make improvements in) all of those factors with the same vigor that now goes into teacher assessment in isolation, there would probably be improvement in education.

References

Albemarle Paper Co. v. Moody, 422 U. S. 405 (1975).

Ayres, Q. W. (1983). Student achievement at predominantly white and predominantly black universities. *American Educational Research Journal, 20,* 291-304.

Bacharach, S. B., Bauer, G., & Shedd, J. (1986). *The learning workplace: The conditions and resources of teaching.* Ithaca, NY: Organizational Analysis and Practice.

Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York: Macmillan.

California Commission on Teacher Education. (1964). *Six areas of teacher competence.* Burlingame, CA: California Teachers Association.

Carini, P. (1975). *Observation and description: An alternative methodology for the investigation of human phenomena.* Grand Forks, ND: University of North Dakota Press.

Connecticut v. Teal, 102 S. Ct. 2525 (1982).

Connell, C. (1987, August 27). 17% of would-be teachers flunk test. *San Francisco Examiner,* p. A-6.

Cross, L. H. (1984, April). *Validation study of the National Teacher Examinations for certification of entry-level teachers in the Commonwealth of Vir-*

ginia. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Cross, L. H. (1987a, April). *Using score statistics from the NTE program tests to make unwarranted inferences about historically black institutions*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Cross, L. H. (1987b, October). *The effectiveness of teacher-training programs at predominantly white and black institutions*. Unpublished paper, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Cross, L. H. (1987c). *NTE programs test score analyses: A summary and interpretation of scores for the period October 1986 through July 1987*. Unpublished paper, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Debra P. v. Turlington, 474 F. Supp. 244 (M. D. Fla. 1979), *aff'd in part, rev'd in part*, 644 F. 2d 347 (5th Cir. 1981), *petition for rehg. and petition for rehg. en banc denied*, 644 F. 2d 397 (5th Cir. 1981).

Dent v. West Virginia, 129 U. S. 114 (1889).

Edelfelt, R. A. (1979). *Inservice teacher education: A concept, an overview*. Syracuse, NY: National Council of States on Inservice Education.

Educator implies test is aimed at black teachers. (1984, April 27). *Washington Post*, pp. C-1, 5.

Elliott, J. (1977). Teachers learn about inquiry/discovery approaches. In R. A. Edelfelt (Ed.), *Inservice education: Criteria for and examples of local programs* (pp. 49-63). Bellingham, WA: Western Washington State College.

Engel, B. (1975). *Handbook on documentation.* Grand Forks, ND: University of North Dakota Press.

Fenstermacher, G. (1978). A philosophical consideration of recent research on teacher effectiveness. In L. Shulman (Ed.), *Review of research in education* (Vol. 6) (pp. 157-185). Itasca, IL: F. E. Peacock.

Few blacks passing Georgia test, study finds. (1983, November 9). *Education Week,* p. 6.

Galvez-Hjornevik, C. (1986, January/February). Mentoring among teachers: A review of the literature. *Journal of Teacher Education,* 6-11.

Gifford, B. R. (1987). Excellence and equity. In L. Rudner et al., *What's happening in teacher testing: An analysis of state teacher testing practices.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Griggs v. Duke Power, 401 U. S. 424 (1971).

Guardians Association v. New York Civil Service Commission, 630 F. 2d 79 (2d Cir. 1980), *cert. denied,* 452 U. S. 940 (1981).

Holly, M. L. (1984). *Keeping a personal-professional journal.* Geelong, Australia: Deakin University Press.

Ingrassi, J. M. (1987, April). *Arizona Teacher Proficiency Examination: Basic skills. Year-end report.* Phoenix, AZ: Arizona Department of Education.

Livingston, S. A., & Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests.* Princeton, NJ: Educational Testing Service.

Medley, D. M. (1982). *Teacher competency testing and*

the teacher educator. Charlottesville, VA: University of Virginia; Reston, VA: Association of Teacher Educators.

Peterson, D. (1986). Florida Performance Measurement System validity studies. In *Teacher evaluation study: Report for 1985-86*. Tampa, FL: University of South Florida, College of Education.

Popkewitz, T. S., Tabachnick, B. R., & Zeichner, J. M. (1979). Dulling the senses: Research in teacher education. *Journal of Teacher Education, 30*, 52-60.

Scheelhaase v. Woodbury Central Community School District, 349 F. Supp. 988 (1972), *rev'd*, 488 F. 2d 237 (1973), *cert. denied*, 94 S. Ct. 3173 (1974).

South Carolina Department of Education. (1987, September). *South Carolina Education Entrance Examination, summary report: October 1984–June 1987*. Columbia, SC: Author.

Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. *Educational Researcher, 13*(9), 11-18.

Superintendent of Public Instruction, Office for Multicultural and Equity Education. (1984). *Selected statistics related to school desegregation on the basis of race and gender: School district staffing*. Olympia, WA: Author.

Tyler v. Vickery, 517 F. 2d 1089 (5th Cir. 1975), *cert. denied*, 426 U. S. 940 (1976).

U. S. v. Nansemond County School Board, 351 F. Supp. 196 (1972), *rev'd*, 492 F. 2d 919 (1974).

U. S. v. North Carolina, 400 F. Supp. 343 (E. D. N. C. 1975), *vacated on other grounds*, 425 F. Supp. 789 (E. D. N. C. 1977).

U. S. v. South Carolina, 445 F. Supp. 1094 (D. C. S. C. 1977), *aff'd mem.*, 434 U. S. 1026 (1978).

U.S. Equal Employment Opportunity Commission. *Uniform guidelines on employee selection procedures*. 29 C. F. R. 1607 (1978).

Washington v. Davis, 426 U. S. 229 (1976).

Webster's ninth new collegiate dictionary. (1986). Springfield, MA: Merriam-Webster.

Winne, P. H., & Marx, R. W. (1977). Reconceptualizing research on teaching. *Journal of Educational Psychology, 69*, 668-678.

Appendix

State-Mandated Testing of Prospective Teachers: Test Used and Purpose

Educational Testing Service National Teacher Exam Core Battery,* for Certification

Arkansas
California**
Hawaii
Idaho
Indiana
Kansas***
Kentucky
Louisiana****
Maine
Maryland
Mississippi
Montana
New Jersey**
New Mexico
New York
North Carolina****
Rhode Island
South Dakota
Tennessee
Virginia

Educational Testing Service National Teacher Exam Specialty Area (Subject-Matter) Tests, for Certification

Arkansas
California (in lieu of courses)
Florida
Hawaii
Indiana
Kentucky
Louisiana
Maryland
Mississippi
New Jersey
New Mexico
North Carolina
South Carolina***
South Dakota
Virginia

Data supplied by Educational Testing Service, Berkeley/Bay Area; effective as of November 2, 1987

*Tests of professional knowledge, general knowledge, and communication skills
**General knowledge, in lieu of courses
***Professional knowledge
****Communication skills and general knowledge used in teacher education admissions

Educational Testing Service Pre-Professional Skills Tests (PPST: Reading, Writing, Mathematics), for Certification

Delaware	New Hampshire
Kansas	Minnesota
Nebraska	West Virginia

Educational Testing Service Pre-Professional Skills Tests (PPST: Reading, Writing, Mathematics), for Admission to Teacher Education

Arkansas	Tennessee
Arizona	Texas
Nevada	Wisconsin
Oklahoma	

State-Conceived and/or -Developed Tests

Alabama: English language proficiency, for admission to teacher education

Arizona: basic skills/pedagogy, for certification

Connecticut: basic skills, for admission to teacher education

Florida: basic skills/professional knowledge, for certification

Indiana: tailored subject exams (ETS-developed), for certification

Pennsylvania: basic skills/general knowledge/professional knowledge/subject matter (ETS-developed), for certification

South Carolina: basic skills, for admission to teacher education

Washington: pedagogy (each teacher training institution to develop its own test), as exit examination and for initial certification

Tests Developed by National Evaluation Systems (NES)

Alabama, Georgia, Oklahoma, Texas, West Virginia: subject matter, for certification

California Basic Educational Skills Test (CBEST: Reading, Writing, Mathematics; Operated by ETS for the State of California)

California, for certification

Oregon, for admission to public teacher education institutions and for certification

Other Standardized Tests

Colorado: California Achievement Test, for admission to teacher education

Missouri: SAT/ACT, for admission to teacher education

Washington: Washington Pre-College Test/SAT, for admission to teacher education

West Virginia: ACT composition, speaking, for certification

Instruments for Observational Assessment of Teaching Performance

Arizona
Connecticut (under development)
Florida
Georgia
Kansas (under development)
North Carolina
Oklahoma
Pennsylvania (under development)
Rhode Island
South Carolina
Tennessee
Texas
Virginia
West Virginia

Proposals, Draft Legislation, Plans

Alabama: new certification tests to replace those currently used

Arizona: subject area (or completion of a major), for alternative certification

Arkansas: subject area (or 6 course hours), for continuing certification; Pre-Professional Skills Test, for admission to teacher education

California: revalidation of 15 Specialty Area tests

Colorado: basic skills (test or GPA), written essay, for admission to teacher education; knowledge and performance assessment, for provisional certification

Connecticut: subject area/professional knowledge, for elementary certification; basic skills/subject area, for alternative certification (validation of National Teacher Examination and customized subject-matter tests by RMC Research Corporation); structured interview assessment of beginning teachers

Florida: subject matter (to be developed by eight Florida universities), for initial certification

Illinois: basic skills/subject matter, for initial certification (to be developed by NES)

Indiana: Pre-Professional Skills Test, eight-campus University of Indiana system, for admission to teacher education

Kentucky: National Teacher Examination Educational Administration and Supervision test, for certification of principals

Massachusetts: subject matter/language and communication skills, for certification

Michigan: basic skills/subject matter, for initial certification

Mississippi: Core General Knowledge and Communication Skills, seven nonpublic colleges, for admission to teacher education; National Teacher Examination, for teacher education, program approval

Missouri: basic skills (to be developed by the University of Missouri), for admission to teacher education; competencies to be defined, as "exit test" (for certification)

Nebraska: subject matter, for initial certification

Nevada: subject matter, professional knowledge, for initial certification

New York: validation of additional subject-matter exams

North Carolina: revalidation of Core and Specialty Area tests

North Dakota: assessment as called for by NCATE standards (not state-mandated)

Ohio: basic skills, for admission to teacher education; professional education/general education/subject matter, for certification

Oregon: subject matter, for initial certification; validation of 10 subject exams additional to those already validated

South Carolina: Core Professional Knowledge/four Specialty Area tests (to be validated by IOX), for certification

Tennessee: 16 subject-matter exams, additional to National Teacher Examination Specialty Area tests currently used

Texas: a state-controlled basic skills test with diagnostic features, for admission to teacher education

Utah: basic skills, for admission to teacher education

Wisconsin: subject matter, for certification

Additional copies may be obtained from
Association of Teacher Educators
1900 Association Drive
Reston, Virginia 22091
(703) 620-3110

63